Author’s response to reviews

Title: Economic burden of symptomatic iron deficiency - A Survey among Swiss women

Authors:

Patricia Blank (patricia.blank@unibas.ch)
Yuki Tomonaga (yuki.tomonaga@uzh.ch)
Thomas Szucs (thomas.szucs@unibas.ch)
Matthias Schwenkglenks (matthias.schwenkglenks@uzh.ch)

Version: 3 Date: 19 Jul 2018

Author’s response to reviews:

Reviewer reports:

Michèle Dramaix-Wilmet, Ph.D. (Reviewer 1): This is an interesting paper, clearly presented but I have a few questions (in particular on the methodology) and remarks listed below.

Methods

1. Selection of the persons included in the consumer panels: was the % of non-respondents important?

Author’s response: As in all consumer panels, there is a certain risk of information bias. Despite this, consumer surveys are a very effective and fast way of generating data regarding a specific research question. Normally, the response rate tends to be high, as respondents have opted in to take part in the research. The participation in the panel was voluntary and non-binding.

The screen-outs were as follows (see also Figure 1):

- First contact: 5’301
- accepted adverse event reporting 98%
- First diagnosis iron deficiency 28%
- Symptoms of iron deficiency 23%
- No chronic disease / pregnancy/ operation 19% >> total included sample (n=1010)

On page 12, we added a statement to address this limitation in the discussion. Also, we added a statement to the screen-outs to the results (page 7).
2. Why were Italian speaking women excluded from the ID survey?

Author’s response: In Switzerland there are three main language regions. The German speaking region is the biggest, followed by the French speaking one. If compared to these regions, the Italian speaking part is very small. The Italian speaking part was excluded, as the sample size would have been too small to detect differences with sufficient precision. Furthermore, we did not expect any differences in the responses. We added a statement to the discussion, page 12.

3. How was the sample size determined? On what basis was it computed?

Author’s response: We did not conduct a formal power analysis. However, we worked with a well experienced agency who conducted the surveys. Based on their and our experience, the sample size of 1000 was deemed large enough to be able to answer our research question with sufficient precision.

4. p. 4, line 72: "Within the survey population", does it mean the whole sample from the consumer panels (> 130'000 active)? In which analyses were the weights applied?

Author’s response: The weights were applied to the study sample. We agree that this is not clearly described. Therefore we changed the sentence to “within the study sample” (page 4, lines 74-77).

5. Data collection: were questionnaires completed on line by participants or administrated by telephone by an investigator?

Author’s response: the questionnaire was completed online (as described on page 4, line 70-73).

6. Spearman correlation coefficient: why does it not provide reliable estimate of effect size? Cannot it be interpreted as a Pearson's correlation coefficient using a rule a thumb for interpretation such as for example those listed below?

* <0.3 : "very weak"
* 0.3-<0.5 : "moderate"
* 0.5-0.8 : "strong"
* 0.8-1.0 "very strong"

Such an interpretation has been done for the two correlation coefficients presented p.10. However, one usual assumption to use Spearman's correlation coefficient is that the two variables should be measured on an ordinal or continuous scale (i.e., interval or ratio scale) and it seems that one of the two variables used to compute correlation coefficient p. 10 was dichotomous (presence of misdiagnosis or sick leave). In that case, Wilcoxon - Mann-Whitney
test is usually used, it provides the same P-value and the difference between the medians of the two groups can provide an idea of effect size.

Author’s response: The Spearman correlation coefficient is per se a statistical measure illustrating the strength of a relationship between two variables. As suggested by the reviewer, the correlation may range from very weak to very strong. The MWW test is commonly used to test if two groups have a similar distribution and can provide an idea of the effect size. For the present work, we felt that it was more relevant to emphasize that the correlation between misdiagnosis and mean energy level was low, instead of focusing on small effect sizes.

Results
7. Figure 1: 266 women (18%) with missing information were eliminated, were the characteristics of these women not too different from those of women included in the analyses?

Author’s response: given that some information on their iron deficiency status was missing, we were forced to exclude them from the sample. We are not able to judge, whether these individuals were different from the rest of the sample.

8. p. 7, lines 154-155: "weeks until a medical doctor was consulted": mean and median are very different, so there should be high values for this number of weeks, what's the range for this variable?

Author’s response: The range of this question was wide: between 0 and 1040 weeks. Given the skewed distribution, the mean and median are very different. We added this information to the results (page 7, lines 158-159).

9. p.8, lines 171-174: what's the test used to compare mean number of consultations and mean duration of sick leave? Were these two variables normally distributed? The methodology refers to 3 tests: t, Mann-Whitney and Chi²; where are these tests applied?

Author’s response: Discrete numeric/ continuous variables were analyzed with a t-test (for non-skewed data) or a nonparametric test (e.g. Mann-Whitney U test for skewed data). Bivariate associations of categorical variables were assessed with the chi-squared test. Mean number of consultations and mean duration of sick leave showed a skewed distribution. Therefore, we used a nonparametric test.

We understand your point, but we decided to remain the description of the statistics and results as it is. Otherwise, we would need to add for all p-values the applied test. We think it is more reader-friendly, if we describe in the methods the different tests and how/ when they were applied.

10. p. 9, line 188: what's the range for sensitivity analysis? Is it 20% of 9.5%? How was this value chosen?
Author’s response: In the sensitivity analysis, the annual incidence of ID diagnosis of 9.5% was changed to 7.5% or 11.5%, which correspond to a variation of ca. +/− 20%. This was rather a conservative range. We have added a statement on the discussion (page 12, Lines 273-4)

Discussion
11. p. 11, line 246: "if they are aware of their ID condition or not", this is not clear for me. Can the undiagnosed patients be aware of their ID condition?

Author’s response: We agree that it is not clear. We deleted the final part of the sentence ("regardless if they are aware of their ID condition or not") (Page 11, line 249)

Figures
12. Figure 4: what does it means "ranked impact"? The two most important factors impacting daily living negatively are not at the top or bottom of the graph.

Author’s response: thank you for this input. We agree and changed the figure legend to “Factors impacting quality of life”

Sucheta Mehra (Reviewer 2):
This appears to be an interesting paper and contributes to the literature, elegantly worked out and written. There are few issues that need explanation and elucidation, which would strengthen the paper

1. Line 47- the sentence is a little extreme, there are various nutrients, the deficiency of which is virtually ubiquitous, zinc being an example. Authors should soften the sentence.

Author’s response: we have adapted the sentence to “Symptomatic ID is a nutrient-related disorder which is significantly prevalent in both developing and industrialized countries,...” (see page 3, line 47).

2. Line 54 - the burden on individuals beyond cost of IV therapy should be a consideration, and expounded upon.

Author’s response: Thank you for this point. We agree and have adapted the sentence accordingly (see page 3, lines 54-56).

3. The background would be strengthened with information on the situation in Switzerland - a description of the language representation, the age breakdown and the prevalence of iron deficiency based on other studies.

Author’s response: Switzerland has four official languages: German (spoken by 2/3 of the population), French (spoken by ¼ of the population), Italian (spoken by less than 10% of the population) and Romansh (spoken by about 0.5% of the population). The age breakdown is 19.3% for 0-19 years, 26.1% for 20-39 years, 34.6% for 40-64 years, 13.7% for 65-79 years and 6.3% for 80+ years in the female population [12]. We did not add this data to the methods, as we think that this information does not add much additional information.
We are clearly saying in the method section, that the sample population is not 100% representative to the Swiss population. The prevalence estimation of ID has been already included in the discussion part. Also, we added the age-breakdown of the sample in table 1.

4. Methods - despite much being made of the population-weighting of the dataset, there is no description of the expected distribution versus the distribution of the women in the study. The appropriateness of the method is difficult without this. Information on the screening process would be useful - were women simply asked if they received an ID diagnosis, was there more to the process?

Author’s response: We have clearly indicated in the methods, that our sample is not 100% representative of the Swiss population (see also point 3). The screen-outs was as follows (see also Figure 1):

- First contact: 5'301: 100
- accepted adverse event reporting: 98% (here, the individuals were asked, if they agree to report any adverse events, 2% did not accept this term)
- First diagnosis iron deficiency: 28% (here, the individuals were asked if they had a first diagnosis of ID within the last 2 years)
- Symptoms of iron deficiency: 23% (this was the question related to any specific symptoms of ID. If they had no ID specific symptoms, they were excluded)
- No chronic disease / pregnancy/ operation: 19% (this was the question if they are suffering from any chronic disease, if they were pregnant or if they had an operation in the last 2 years)
- >> total included sample (n=1010)

We have added a statement in the method section (page 7).

5. Line 91 - why is no IRB permission needed for this sort of data collection? If identifiers were not collected, this needs to be stated clearly.

Author’s response: The participation in the panel is voluntary and non-binding. At the beginning of the survey, participants' consent to take part was explicitly asked. No identifiers were collected. Therefore, no specific ethical approval was required for this type of research. We have added this information to the method section (page 5, line 95).

6. Line 101 - the whole paragraph describes Spearman's correlation, but the data presented in the tables does not show any correlational information. This paragraph does not add any value to the paper.

Author’s response: Associations between energy level (before ID diagnosis) and present misdiagnosis or sick leave were assessed by Spearman's correlation coefficient. We have adapted
this on page 5. We would like to leave the description in the method section. We added the description of the variables to make clear why we used Spearman's correlation (Methods, page 5).

7. Line 128 - It would be useful to describe the approach, considering the paper being referred to is not in English.

Author’s response: We added as reference a recently published study describing the costs adaptation in English language. We did not add any additional explanation of the approach, as we think this would blow up the method section.

8. Line 152 and table 1 - age categories would be useful to show.

Author’s response: As reported in Table 1, mean age was 33.5 years, with a 95% CI from 32.9 to 34.1 years. We added also additional age-categories and the range of the sample (in years).

9. Line 271 - It is also assumed that women who have ID seek treatment for it, this should also be elucidated, not merely assumed that all patients be treated.

Author’s response: It could be assumed that all patients with ID consulting a medical doctor would be treated. However, this might lead to an overestimation of the costs, since in reality not all patients would wish or receive a medical treatment. In our analysis, we applied only costs to those patients receiving a therapy, according to our study population (i.e. around 50% of the misdiagnosed patients). This should have reduced the overestimation, at least in part. In the discussion, there is one paragraph elucidating this assumption. We however re-formulated the sentence to make it clearer (Discussion, page 12).